



University of Pennsylvania
ScholarlyCommons

Departmental Papers (Chemistry)

Department of Chemistry

5-8-2012

Computational Design of a Protein Crystal

Christopher J Lanci
University of Pennsylvania

Christopher M MacDermaid
University of Pennsylvania

Seung-gu Keng
University of Pennsylvania

Rudresh Acharya

Benjamin North

See next page for additional authors

Follow this and additional works at: http://repository.upenn.edu/chemistry_papers

 Part of the [Biochemistry Commons](#), [Organic Chemistry Commons](#), and the [Physical Chemistry Commons](#)

Recommended Citation

Lanci, C., MacDermaid, C., Keng, S., Acharya, R., North, B., Yang, X., Qiu, X. J., DeGrado, W. F., & Saven, J. G. (2012). Computational Design of a Protein Crystal. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (19), 7304-7309. <http://dx.doi.org/10.1073/pnas.1112595109>

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/chemistry_papers/22
For more information, please contact repository@pobox.upenn.edu.

Computational Design of a Protein Crystal

Abstract

Protein crystals have catalytic and materials applications and are central to efforts in structural biology and therapeutic development. Designing predetermined crystal structures can be subtle given the complexity of proteins and the noncovalent interactions that govern crystallization. De novo protein design provides an approach to engineer highly complex nanoscale molecular structures, and often the positions of atoms can be programmed with sub-Å precision. Herein, a computational approach is presented for the design of proteins that self-assemble in three dimensions to yield macroscopic crystals. A three-helix coiled-coil protein is designed de novo to form a polar, layered, three-dimensional crystal having the P6 space group, which has a “honeycomb-like” structure and hexameric channels that span the crystal. The approach involves: (i) creating an ensemble of crystalline structures consistent with the targeted symmetry; (ii) characterizing this ensemble to identify “designable” structures from minima in the sequence-structure energy landscape and designing sequences for these structures; (iii) experimentally characterizing candidate proteins. A 2.1 Å resolution X-ray crystal structure of one such designed protein exhibits sub-Å agreement [backbone root mean square deviation (rmsd)] with the computational model of the crystal. This approach to crystal design has potential applications to the de novo design of nanostructured materials and to the modification of natural proteins to facilitate X-ray crystallographic analysis.

Keywords

biomaterials, computational protein design, crystal engineering, protein crystallization, self-assembly

Disciplines

Biochemistry | Organic Chemistry | Physical Chemistry

Author(s)

Christopher J Lanci, Christopher M MacDermaid, Seung-gu Keng, Rudresh Acharya, Benjamin North, Xi Yang, X Jade Qiu, William F. DeGrado, and Jeffery G. Saven

Computational design of a protein crystal

Christopher J. Lanci^{a,1}, Christopher M. MacDermaid^{a,1}, Seung-gu Kang^a, Rudresh Acharya^b, Benjamin North^b, Xi Yang^a, X. Jade Qiu^{b,c}, William F. DeGrado^{b,c}, and Jeffery G. Saven^{a,2}

^aDepartment of Chemistry, University of Pennsylvania, Philadelphia, PA 19104-6323; ^bDepartment of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6059; and ^cDepartment of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158-9001

Edited by David Eisenberg, University of California, Los Angeles, CA, and approved March 8, 2012 (received for review August 2, 2011)

Protein crystals have catalytic and materials applications and are central to efforts in structural biology and therapeutic development. Designing predetermined crystal structures can be subtle given the complexity of proteins and the noncovalent interactions that govern crystallization. De novo protein design provides an approach to engineer highly complex nanoscale molecular structures, and often the positions of atoms can be programmed with sub-Å precision. Herein, a computational approach is presented for the design of proteins that self-assemble in three dimensions to yield macroscopic crystals. A three-helix coiled-coil protein is designed de novo to form a polar, layered, three-dimensional crystal having the P6 space group, which has a “honeycomb-like” structure and hexameric channels that span the crystal. The approach involves: (i) creating an ensemble of crystalline structures consistent with the targeted symmetry; (ii) characterizing this ensemble to identify “designable” structures from minima in the sequence-structure energy landscape and designing sequences for these structures; (iii) experimentally characterizing candidate proteins. A 2.1 Å resolution X-ray crystal structure of one such designed protein exhibits sub-Å agreement [backbone root mean square deviation (rmsd)] with the computational model of the crystal. This approach to crystal design has potential applications to the de novo design of nanostructured materials and to the modification of natural proteins to facilitate X-ray crystallographic analysis.

biomaterials | computational protein design | crystal engineering | protein crystallization | self-assembly

Molecular design provides powerful tools for exploring how molecular properties dictate macroscopic structure and function. One of the most precise forms of self-organization, crystallization achieves orientation and symmetry across many length scales and can be leveraged to engineer materials with well-defined molecular order (1). In addition to their central role in structure determination, molecular crystals have many applications, including nanoparticle templating (2, 3), nonlinear optical devices (4), molecular scaffolding (5), and porous frameworks (6). A predictive understanding of how to achieve self-assembled macroscale structure and desired properties remains challenging, however, particularly when large, conformationally flexible molecules are employed.

Small synthetic molecules have been designed that display complementary functional groups in a manner consistent with a chosen crystal structure. Intermolecular contacts are often patterned using strong, directional interactions (e.g., hydrogen bonding, metal-coordination, and electrostatic interactions) (7). The constituent molecules, however, are typically small and synthetic, thus limiting size, shape, and functionality. Hence, simultaneously achieving functional properties and the presentation of complementary intermolecular interactions that confer a targeted crystal structure can be difficult. Commonly, weak intermolecular forces stabilize crystalline ordering, and as a result, quantitative, predictive approaches to crystal design are challenging even with small molecules (7, 8). In addition, within most crystals, molecules align in a nonpolar fashion (9), but polar crystalline ordering is desired for many applications, e.g., nonlinear optical materials (6).

Crystals of proteins and other biomacromolecules are widely studied, and they play a crucial role in structural biology, where realization of diffraction-quality crystals is often the bottleneck of structural genomics efforts (10, 11). Protein and biopolymer crystals also have wide applicability to catalysis, therapeutic formulation, and biomaterials (12). However, the space group and local crystalline structure of biopolymer crystals are most often not at the control of the researcher and are only identified after extensive screening and optimization of conditions that yield a diffraction-quality crystal. To facilitate crystallization, variations are often introduced into proteins, with the goals of improving solubility (13), reducing conformational variability of amino acid side chains (14, 15), creating synthetic, high-symmetry oligomers (16, 17), and making use of known motifs at protein interfaces, e.g., an interhelical leucine zipper (18). Alternatively, ab initio protein crystal design has the potential to alleviate difficulties associated with protein crystallization and structure determination.

The size, flexibility, and complexity of proteins impede the design of sequences that form a particular crystal lattice. In some cases, intermolecular interactions can be programmed using structures and rules gathered from natural systems. Biopolymer crystals of predetermined structure have been designed using DNA (19). Nanometer scale nanohedra and filaments have been prespecified using heterodimers of known self-associating protein domains (20). The structures and functions of these previously studied self-assembled systems are limited, however, to those achievable with DNA hybridization or preexisting protein domains. Much greater versatility in structure and functionality is accessible via the de novo design of protein-protein interfaces within a targeted crystal structure. Such interfaces, stabilized by noncovalent interactions, could be designed into a wide variety of structures via careful selection of sequence and a given crystalline lattice.

The design of proteins that self-assemble in three-dimensions would enable the creation of crystals having predetermined molecular structures. Such a technology has many potential applications, and the current investigation targeted protein crystals containing aqueous channels that might allow future positioning of cofactors and guest molecules. For many optical applications (e.g., nonlinear optics), it is also important to create “polar” arrays that specify the unidirectional orientation of chromophores. In this work, a layered, porous P6 space group (Fig. 1A) offers desirable features: (i) high symmetry, which can facilitate data collection and structure determination using X-ray crystallogra-

Author contributions: C.J.L., C.M.M., B.N., X.Y., W.F.D., and J.G.S. designed research; C.J.L., C.M.M., S.-g.K., R.A., B.N., X.Y., and X.J.Q. performed research; C.J.L., C.M.M., S.-g.K., R.A., B.N., X.Y., X.J.Q., W.F.D., and J.G.S. analyzed data; and C.J.L., C.M.M., R.A., B.N., W.F.D., and J.G.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org [PDB ID codes 3V86 (P6-a) and 4DAC (P6-d)].

¹C.J.L. and C.M.M. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: saven@sas.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112595109/-DCSupplemental.

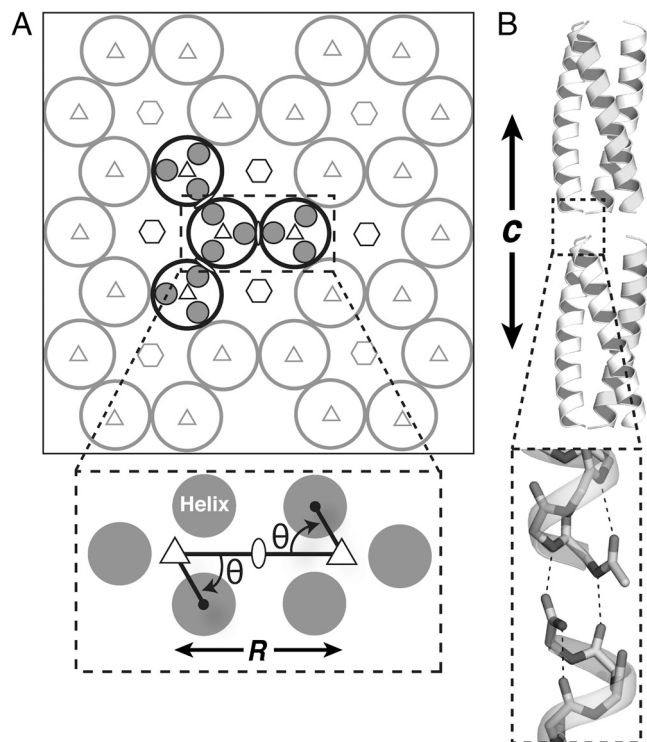


Fig. 1. (A) One layer of the P6 crystal viewed along the unit cell *c*-axis. Protein (open circle) comprises three identical helices (small filled circles). *R* is the intralayer distance between neighboring proteins. θ is the rotation angle about the protein's superhelical axis. C_N symmetric rotation axes (*N*-sided polygons) and C_2 axes (ovals) indicated. (B) Two adjacent layers of the P6 crystal and helical hydrogen bonding at the interlayer interface.

phy, (ii) tubular solvent channels extending through the crystal, and (iii) polarity, i.e., large-scale parallel orientation of the proteins. Moreover, the P6 space group is one of the less common (2), occurring in only 0.1% of known protein crystal structures (21). Thus, achieving a designed crystal with P6 ordering is a stringent test of the approach.

Computational protein design continues to advance (22–31), and herein a strategy for designing a predetermined protein crystal structure is presented. A three-helix coiled-coil protein was designed de novo to form a polar, layered three-dimensional crystal having the P6 space group. Candidate proteins were identified from local energy minima on a sequence-structure energy landscape. X-ray crystallographic studies revealed that one such designed protein is in subangstrom (C_{α} rmsd < 0.70 Å) agreement with the computationally designed model.

Results and Discussion

Computational Design. In many applications of protein design, designability of a structure is made accessible by using structures from naturally occurring proteins (24, 27, 32, 33), but here both the structure of the protein as well as its crystalline ordering are specified de novo. A trimeric coiled-coil protein is designed to fold into a stable unit that then further assembles into a crystalline structure. Related hierarchical strategies have been employed to design proteins and their higher order assemblies (34, 35). The design of proteins for a chosen crystalline structure is divided into (i) identification of the constituent protein, (ii) determination of a set of physically accessible, three-dimensional, crystalline arrays of the protein backbone, (iii) identification of candidate crystalline structures within this set, (iv) design of sequences for these structures, and (v) experimental synthesis, screening for crystallization conditions, and structure elucidation.

The target protein was a mathematically created (idealized) homotrimeric parallel coiled-coil protein having C_3 symmetry, a superhelical pitch of 120 Å, and (initially) 27 residues per helix (36). Similar structures have been observed in prototypical three-stranded coiled coils (37–39). The 8 interior (*a* and *d* heptad) positions of each helix contained well-packed Val and Leu residues (39). At the remaining exterior positions, all amino acids except cysteine and proline were considered in the design of crystalline ordering.

Within a crystal of P6 symmetry, many symmetry-related configurations involving the multiple copies of the protein are possible. Choosing a high-symmetry space group reduces the number of degrees of freedom D available to the protein. In general, protein units within P6 have $D = 5$ (40), but here D is reduced to $D = 2$ when generating the crystalline configurations: (i) The C_3 axis of the protein is chosen to coincide with a C_3 axis in the crystal (Fig. 1A), and (ii) the length of the c -unit cell vector is fixed at $c = 40.7 \text{ \AA}$ to achieve backbone helical hydrogen bonding between helix termini (Fig. 1B). Each helix's approximately 28 residue equivalents displace the N terminus approximately 120° (one-third turn) about the protein's superhelical C_3 axis relative to the C terminus, resulting in a pseudocontiguous super helix and hydrophobic core. Complementary hydrogen bonding and hydrophobic interactions are specified at the interlayer interface. Candidate crystalline arrays consistent with P6 were generated by varying intralayer degrees of freedom R and θ . R is the distance between neighboring proteins ($a = b = 3^{1/2}R$). θ is the angle of rotation of each protein about its superhelical C_3 axis (Fig. 1A), and this rotation maintains the C_2 axis between nearest neighbors. N-terminal acetylated proteins were used to characterize crystalline arrays of structures energetically. All sequence design calculations were performed in the context of the local 3-D crystalline array.

In a first attempt, a minimum energy crystalline configuration was identified based upon backbone interactions on the protein. “Minimal side chain” calculations were used where Gly was modeled at each of the protein’s exterior positions. A grid search was performed over R and θ ($15 \text{ \AA} < R < 22 \text{ \AA}$, increment $\Delta R = 0.2 \text{ \AA}$; $0 < \theta < 60^\circ$, increment $\Delta\theta = 5^\circ$, 520 structures), which identified the lowest energy configuration ($R = 19.0 \text{ \AA}$, $\theta = 0.0^\circ$) [AMBER (41)]. Subsequently, a sequence was computationally designed using this structure. In the local crystalline environment (42), the exterior residues of each helix (19/27) were computationally designed consistent with the periodic symmetry (42, 43).

Experimental Characterization of Protein *P6-a*. A single low-energy sequence for the identified structure, *P6-a*, was selected, synthesized and formed diffraction-quality crystals overnight at room temperature using a standard crystal screen. *P6-a* crystallized in the apolar P321 space group, and the structure was solved to 2.9 Å resolution. The crystal contained columnar, hexagonal pores resembling the target, but neighboring proteins were antiparallel (Fig. 24). A model of the *P6-a* sequence in the observed P321 crystalline array was built that contains the most probable side-chain conformations. For the crystallographic structure and this model, the computed interaction energy per protein within the P321 structure was above that of the P6 structure, i.e., the calculation did not discern P321 as the preferred crystalline structure. The observed frequency of P321, however, is more than three times the frequency of P6 (21, 40). In addition, P321 can accommodate deviations from the planar layering present in P6. The formation of P321 by this protein may be kinetically and entropically more facile than that of P6. Furthermore, the antiparallel packing involves an extended right-handed “glycine zipper” motif (GX₃GX₃A) (44), which is similar to the GX₃G motif but is found in both parallel and antiparallel orientations (45, 46).

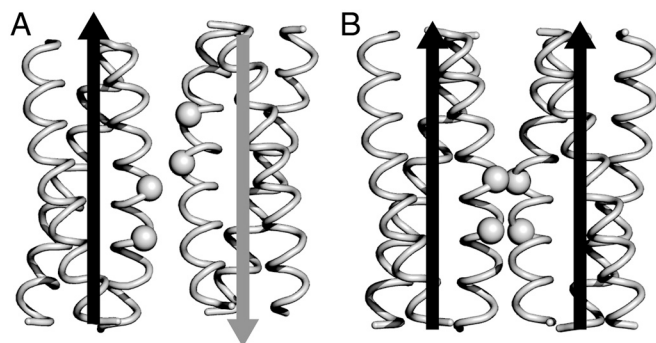


Fig. 2. (A) Crystallographic structure of *P6-a* has P321 symmetry; neighboring proteins are antiparallel. (B) *P6-a* model ($R = 19.0$ Å and $\theta = 0.0^\circ$). Gly residues indicated as spheres. Arrows are directed from N to C terminus.

Sequence-Structure Energy Landscape. Given the subtlety of engineering proteins consistent with the targeted polar crystal, the scope of the design was broadened to characterize a wide range of structures and sequences concomitantly, i.e., to survey the sequence-structure designability landscape of proteins consistent with P6 symmetry. The protein was shortened to 26 residues (28 counting the N-terminal acetyl and the C-terminal amide (CO-NH₂) groups as residues) (39), in order to permit a one-amide (1.5 Å rise/residue) gap between helix termini that ameliorates possible steric interactions and allows facile readjustment of the individual coiled-coil proteins upon crystallization. A higher resolution sampling of the sequence-structure energy landscape was undertaken ($15 \text{ Å} < R < 22 \text{ Å}$, $\Delta R = 0.1 \text{ Å}$; $0 < \theta < 120^\circ$, $\Delta\theta = 0.5^\circ$; 19,200 structures) to allow greater possibility for identifying proteins specific for P6 symmetry. R and θ values, where the backbone atoms overlap, were discounted. In the context of a protein's nearest neighbors within each crystalline configuration, the site-specific probabilities of the 18 allowed amino acids, their side-chain conformations, and the weighted average energy over these sequence-rotamer states $E(R, \theta)$ were estimated computationally (30, 31, 47). Effective solvation energy functions were not used (31), and no constraint was placed on the net charge of the protein. A symmetry assumption was applied, where equivalent residues on all helices had the same identities and same rotamer states (42). This resulted in variation of only 18 unique, exterior residues. $E(R, \theta)$ captures the contributions of predominantly low-energy sequences (31, 48), and it yields a sequence-structure energy landscape (Fig. 3). Minima identify candidate crystalline structures that potentially support such low-energy sequences. In arriving at specific sequences for given R and θ , the calculations proceeded iteratively. With each iteration, the most probable amino acid was selected at sites where its probability was at least twice the next most probable. In the final iteration, a helix propensity (49) constraint was imposed and chosen to have a value consistent with parallel homotrimeric coiled-coil proteins of similar size. The final sequence contained the most probable amino acids at the remaining variable residues.

Experimental Sampling of the Landscape. Four low-energy sequences (*P6-b*, *P6-c*, *P6-d*, *P6-e*; Fig. 3C), which sample distinct energy landscape minima (Fig. 3A) and possess sequence properties consistent with coiled-coil proteins, were selected for synthesis and purification. Proteins *P6-c* and *P6-d* formed crystals overnight, which grew to a size of approximately 0.2 mm in one week at room temperature; Proteins *P6-b* and *P6-e* did not crystallize using the same screens and conditions. The *P6-c* crystals did not diffract well enough for structure determination. Analysis of X-ray diffraction data from the *P6-d* crystal obtained using an in-house diffractometer revealed a high-resolution (2.1 Å) structure consistent with the targeted P6 crystalline lattice (Fig. 4). In solution, *P6-d* increased in helical content with increasing peptide

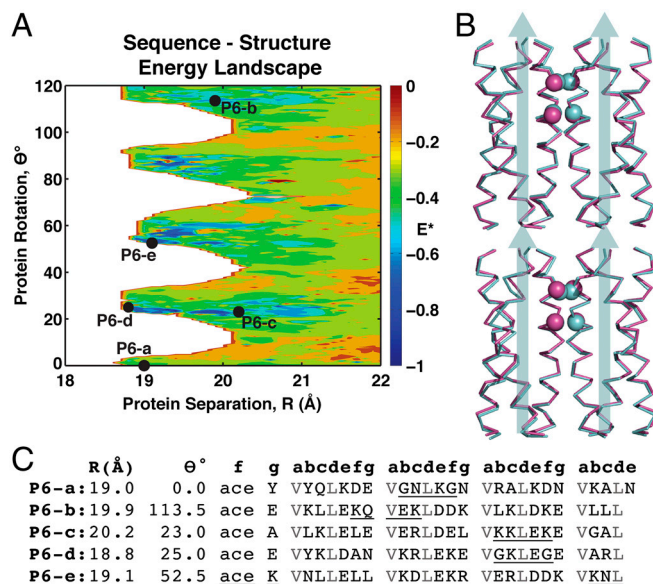


Fig. 3. (A) Sequence-structure energy landscape as a function of R and θ , $E^* = (E - E_{R=\infty}) / (E_{\min} - E_{R=\infty})$ where E is an internal energy over sequences and side-chain conformations for each structure, and E_{\min} is the minimum of E . White region is disallowed due to backbone van der Waals overlap. Synthesized sequences are indicated. (B) Crystallographic (cyan) and computational model (magenta) structures (α atoms) of four neighboring proteins within the *P6-d* crystal; RMSD for all backbone heavy atoms is 0.73 Å. α atoms of Gly residues are highlighted as spheres. (C) Sequences of designed proteins. The segment at the closest point of approach between the neighboring proteins is underlined. The GX₃G motif is displaced one heptad nearer the C terminus in *P6-d* than *P6-a*. Interior residues are gray; "ace" denotes acetylated N terminus.

concentration and was highly helical near concentrations used for crystallization, consistent with the folding of the trimeric protein in solution and subsequent crystallization.

Comparison Between the Model and Crystallographic Structures. The *P6-d* proteins are oriented in the targeted all-parallel polar arrangement with $R = 18.4$ Å, in agreement with the template model ($R = 18.8$ Å). Interestingly, the crystal structure of protein *P6-d* has four helices (ABCD) in the asymmetric unit, which differ mostly in side-chain conformations, yielding two distinct protein structures: *P6-d.1*, comprising helices A, B, and C within the asymmetric unit and *P6-d.2*, a C₃ symmetric protein comprising only helix D. The two protein structures are nearly identical with a backbone rmsd = 0.45 Å. The backbone structures of *P6-d* are in excellent agreement with the computational template [C_{α} rmsd = 0.40 Å (*P6-d.2*) and 0.61 Å (*P6-d.1*)]. Close agreement is also observed when a pair of dimers comprising adjacent proteins in the crystal are aligned and compared to the computational template. A pair of laterally neighboring proteins (*P6-d.1*—*P6-d.2*) has backbone rmsd = 0.64 Å, and an axially neighboring pair (*P6-d.1*—*P6-d.1*) has backbone rmsd = 0.70 Å (Fig. 3B). A comparison of peptides within the asymmetric unit to their coordinates in the model yields rmsd = 1.3 Å (all resolvable atoms) and rmsd = 0.68 Å (backbone only) (Fig. 5A).

Structural Characterization. Although the energy landscape strategy allows many possible interprotein separations and contact-mediating residues, in the *P6-d* crystal, backbone van der Waals contacts are present at the intralayer point of closest approach between proteins (6.4 Å between alpha helical axes). At this interface, the design yields a parallel GX₃G motif, often observed at interhelical contacts (45, 46, 50, 51). As discussed above, *P6-a* has a related antiparallel GX₃GX₃A motif, which may be expected because the single template crystalline structure was iden-

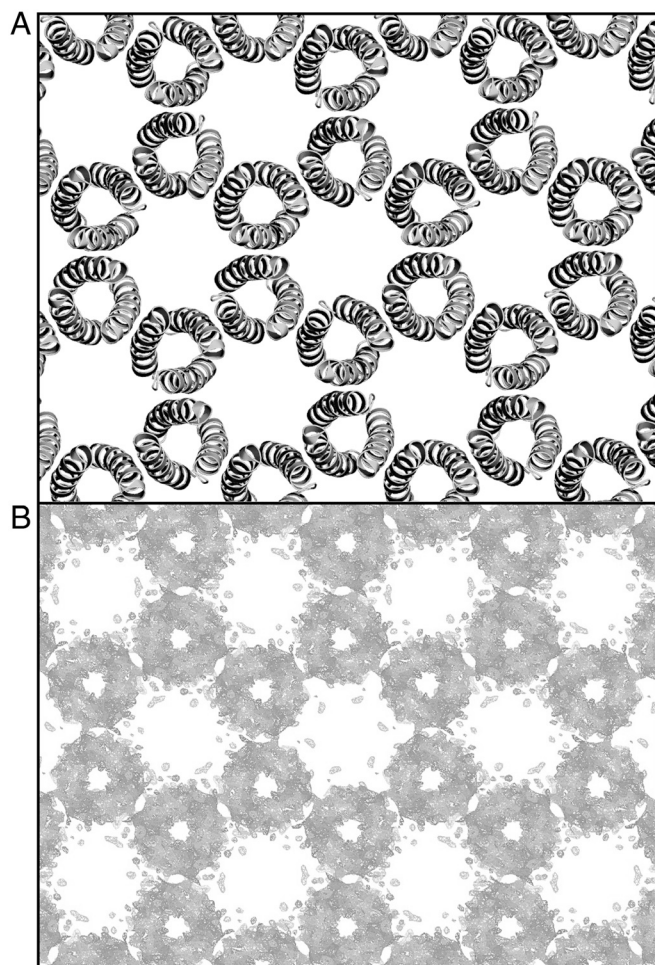


Fig. 4. (A) One layer of *P6-d* model crystalline structure. (B) Resolvable electron density of *P6-d* crystal: omit map ($2F_o - F_c$) contoured at 2σ .

tified using a model having a glycine exterior. *P6-b*, *P6-c*, and *P6-e* do not have this motif. Glycine and small amino acids have been suggested to promote crystallization through the reduction of surface entropy (8, 14, 52–54), and the presence of GX₃G motifs at each protein-protein interface is consistent with a tightly packed crystal. Furthermore the volume per molecular weight (Matthew's coefficient) (55) $V_m = 1.99 \text{ \AA}^3/\text{Da}$ is considerably

lower than that typically observed for protein crystals ($\langle V_m \rangle = 2.68 \text{ \AA}^3/\text{Da}$, $1.5 \text{ \AA}^3/\text{Da} < V_m < 5 \text{ \AA}^3/\text{Da}$) (21). The solvent content (36%) is less than that observed for typical protein crystals (51%), hexagonal protein crystals (57%), and protein crystals with four polypeptide chains in the asymmetric unit (51%) (21, 55, 56).

Backbone and side-chain interactions observed in the crystal structure are consistent with the parallel protein-protein orientation (Fig. 5 *B* and *C*). The C-terminal location of the GX₃G motif in *P6-d* imposes asymmetry along the helix favoring parallel orientations, whereas in *P6-a*, this motif is near the midpoint of the protein (G₁₀X₃G₁₄), and approximately 180° rotation about this point of contact permits antiparallel packing (Fig. 2 *A* and *B*). The large residue (Tyr1) at the N terminus of *P6-a* may also contribute to destabilization of the parallel configuration. Charged side-chains often have high conformational entropy, and they are believed to disrupt crystal-packing interfaces (8, 52, 53, 57, 58). These residues are frequently mutated to induce crystallographic order (54, 59). Lys, in particular, appears infrequently at such interfaces (52, 53). Nevertheless, the *P6-d* structure provides evidence that ionizable residues such as Lys, Arg and Asp can form well-defined, complementary interactions in a designed, high-density crystal (Fig. 5*B*).

The *P6-d* structure suggests that well-defined crystal contacts likely require low conformational entropy amino acids to drive crystallization, e.g. GX₃G motif, augmented by complementary side-chain interactions (54, 59). Compared to *P6-d*, the models of *P6-b*, *P6-c* and *P6-e* have larger interprotein separations, *R* = 19.9 Å, 20.2 Å, and 19.1 Å, respectively, and do not contain the “small-X₃-small” motif. Larger residues may be incorporated at the crystal contact positions, which may decrease the propensity toward crystalline ordering. In the model of *P6-e*, the point of closest approach is at the C terminus and the complementary backbone-backbone contact area is necessarily reduced compared to *P6-d*.

Conclusion. *P6-d* forms the first de novo designed protein crystal. The high-symmetry, high-density P6 structure possesses an uncommon polar arrangement throughout (6, 9), providing a route to controlled supramolecular parallel alignment of proteins (1, 6). De novo protein crystal design presents unique challenges. The high degree of self-assembly required to achieve a targeted crystal structure can also lead to aggregation and poor solubility. The probabilistic computational protein design methodology, however, provides a unique view of the sequence-structure energy landscape compatible with a chosen crystal lattice. Candidate

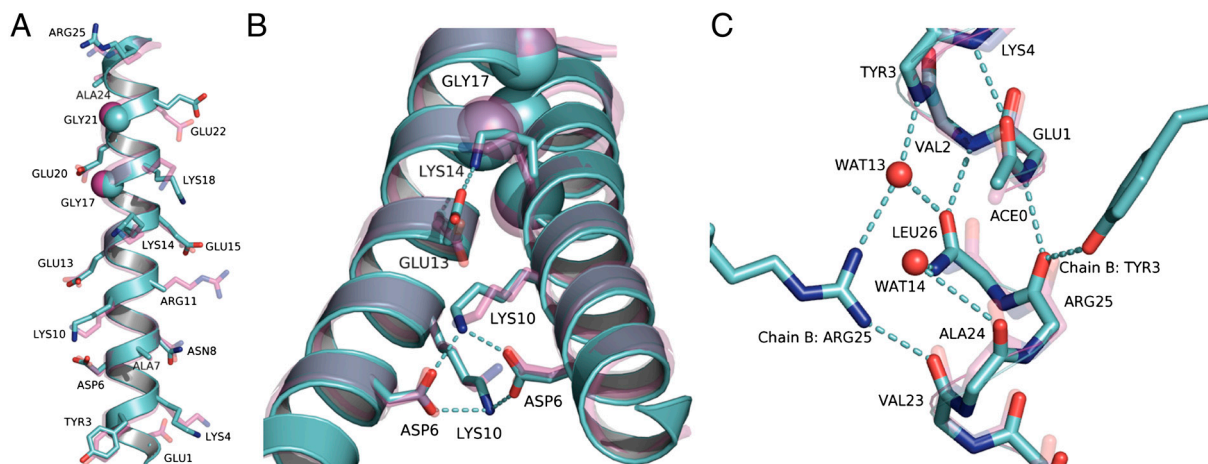


Fig. 5. Computational model (magenta) and crystal structure (cyan). (A) Helix C having an all atom RMSD of 1.0 Å. Val and Leu side chains are not rendered for clarity. Flexible Arg₁₁ exhibits no electron density. (B) Helices C and D involved in interprotein crystal contact. Ca atoms of Gly₁₇ and Gly₂₁ (spheres). Designed complementary electrostatic interactions, Asp₆-Lys₁₀ and Glu₁₃-Lys₁₄ span the gap between the helices. (C) Hydrogen bonding across the interlayer interface.

sequences and structures, which need not be nearby in either structure or sequence space, can be identified using an “aerial view” of the sequence-structure energy landscape, increasing the likelihood of identifying a protein that forms the targeted crystal structure. Computational protein crystal design can be of great utility to structural biology and genomics, where protein crystallization is essential to obtain high-resolution structures (10, 11). Partial (low-resolution) structural information or comparative models may be used in building model crystals and designing crystal contacts, potentially resulting in sequences with improved crystal quality and higher resolution structural information. Efforts in crystal design can further our understanding of the effects of mutation and modifying crystallization conditions. Targeting high-symmetry space groups and reducing crystal solvent content can improve the quality of X-ray diffraction data and simplify structure determination. Given the wide array of functionalities and cofactors that can be incorporated into proteins, targeted protein crystal design can also provide a vehicle to explore new protein-based materials and nanostructures.

Materials and Methods

Computational Design. The structure of the protein and the crystalline configurations were generated by applying the appropriate symmetry operations (translations and rotations) consistent with the C_3 symmetry of the protein and the P6 space group. R is related to the unit cell parameters $a = b = \sqrt{3}R$, and the distance between the centers of mass of two interfacing proteins in adjacent layers is dictated by the c unit cell parameter. The rotational orientation between two neighboring intralayer proteins (θ) is defined in the plane perpendicular to the C_3 axis of a protein. Due to the protein's C_3 symmetry, only the range $0^\circ \leq \theta \leq 120^\circ$ is unique. The angle θ is varied by rotating the proteins about their C_3 axes such that the C_2 symmetry axis between adjacent proteins (and P6 symmetry) was retained. Variation of R and θ yields an ensemble of candidate crystalline structures. Structures with overlapping atoms or other high-energy interactions were filtered using AMBER (41) or CHARMM27 (60, 61) potentials, e.g., structures with $R < 18.0$ Å possess overlapping backbone atoms. In arriving at the sequence-structure energy landscape (Fig. 3A), a statistical thermodynamic method for calculating site-specific probabilities of the amino acids and the average energy over sequences was applied (30, 31). The method uses

the AMBER energy function (41) and a discrete rotamer library (62). Iterative calculations, wherein the most probable amino acids are specified after each iteration, are used to identify specific sequences within minima on the sequence energy landscape.

Crystal Screening and Structural Determination. For initial crystal screening peptides, P6-b, P6-d, and P6-e were ordered from Genscript (60 mg scale, desalt purity). To obtain larger quantities, sequences P6-a, P6-c, P6-d, and P6-e were synthesized (100 μ mol scale) via solid phase peptide synthesis, using 9-fluorenylmethoxycarbonyl (Fmoc), chemistry and upon resin cleavage purified by reverse-phase high-performance liquid chromatography (HPLC). Crystals of P6-a and P6-d were grown at room temperature using the hanging drop vapor diffusion method and flash frozen in liquid nitrogen prior to data collection. Multiple crystals of peptide P6-a were obtained with a peptide concentration of 10 mg/mL with reservoir solution (0.01 M cobalt (II) chloride hexahydrate, 0.1 M MES monohydrate pH 6.5, 1.8 M ammonium sulfate). A single crystal of peptide P6-d was obtained with a peptide concentration of 7.2 mg/mL with reservoir solution (0.17 M ammonium acetate, 0.085 M tri-sodium citrate dihydrate pH 5.6, 25.5% vol/vol PEG 4000, 15% vol/vol glycerol). X-ray diffraction data were collected using a Rigaku R-Axis IV image plate detector equipped with a Cu (K_α) radiation source, and the structures were solved by molecular replacement. For protein P6-a, a poly-alanine model generated from coordinates of a single helix from a similar protein (PDB ID code 1COI) (38), was used as a search model. The refinement strategy converged the $R_{\text{work}}/R_{\text{free}} = 0.175/0.216$. For protein P6-d, four helices from the computational design model served as the initial search ensemble and were used concurrently. The refinement strategy converged $R_{\text{work}}/R_{\text{free}} = 0.148/0.205$.

ACKNOWLEDGMENTS. We thank Axel Kohlmeyer for assistance with modeling (Topo Tools); David Christianson, Dan Dowling, Julie Aaron, Pat Lombardi, Mustafa Koksai, and Steven E. Stayrook for their assistance with structure determination and refinement; E. James Peterson, Ivan Dmochowski, Jacob Goldberg and Glen Liszczak for their assistance with instrumentation and synthesis; and Ilan Samish for helpful comments. The authors acknowledge support from the Penn NanoBio Interface Center (National Science Foundation NSEC DMR-0425780), the US Department of Energy (DE-FG02-04ER46156), the National Institutes of Health (HL085303, GM54616), and the University of Pennsylvania's Laboratory for Research on the Structure of Matter (NSF MRSEC DMR05-20020) for infrastructural support.

- Braga D, Grepioni F, Desiraju G (1998) Crystal engineering and organometallic architecture. *Chem Rev* 98:1375–1405.
- Yeates TO, Padilla JE (2002) Designing supramolecular protein assemblies. *Curr Opin Struct Biol* 12:464–470.
- Sleytr UB, Sara M, Pum D, Schuster B (2001) Characterization and use of crystalline bacterial cell surface layers. *Prog Surf Sci* 68:231–278.
- Hollingsworth MD (2002) Crystal engineering: from structure to function. *Science* 295:2410–2413.
- Seeman NC (1982) Nucleic acid junctions and lattices. *J Theor Biol* 99:237–247.
- Holman KT, Pivovar AM, Ward MD (2001) Engineering crystal symmetry and polar order in molecular host frameworks. *Science* 294:1907–1911.
- Desiraju GR (2007) Crystal engineering: A holistic view. *Angewandte Chemie-International Edition* 46:8342–8356.
- Price WN, 2nd, et al. (2009) Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotechnol* 27:51–57.
- Glaser R (2007) Polar order by rational design: Crystal engineering with parallel befoamphiphile monolayers. *Acc Chem Res* 40:9–17.
- Terwilliger TC, Stuart D, Yokoyama S (2009) Lessons from Structural Genomics. *Annu Rev Biophys* 38:371–383.
- Chandonia JM, Brenner SE (2006) The impact of structural genomics: Expectations and outcomes. *Science* 311:347–351.
- Margolin AL, Navia MA (2001) Protein crystals as novel catalytic materials. *Angewandte Chemie-International Edition* 40:2204–2222.
- Derewenda ZS (2010) Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Crystallogr D Biol Crystallogr* 66(Pt 5): 604–615.
- Cooper D, et al. (2007) Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr D Biol Crystallogr* 63(Pt 5):636–645.
- Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D (2007) Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci* 16:1569–1576.
- Banatao DR, et al. (2006) An approach to crystallizing proteins by synthetic symmetrization. *Proc Natl Acad Sci USA* 103:16230–16235.
- Andre I, Strauss CEM, Kaplan DB, Bradley P, Baker D (2008) Emergence of symmetry in homooligomeric biological assemblies. *Proc Natl Acad Sci USA* 105:16148–16152.
- Yamada H, et al. (2007) 'Crystal lattice engineering', an approach to engineer protein crystal contacts by creating intermolecular symmetry: crystallization and structure determination of a mutant human RNase 1 with a hydrophobic interface of leucines. *Protein Sci* 16:1389–1397.
- Zheng J, et al. (2009) From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* 461:74–77.
- Padilla JE, Colovos C, Yeates TO (2001) Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci USA* 98:2217–2221.
- Chruszcz M, et al. (2008) Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? *Protein Sci* 17:623–632.
- Kortemme T, Baker D (2004) Computational design of protein-protein interactions. *Curr Opin Chem Biol* 8:91–97.
- Kang SG, Saven JG (2007) Computational protein design: structure, function and combinatorial diversity. *Curr Opin Chem Biol* 11:329–334.
- Grigoryan G, Reinke AW, Keating AE (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* 458:859–U852.
- Jha RK, et al. (2010) Computational design of a PAK1 binding protein. *J Mol Biol* 400:257–270.
- Friedland GD, Kortemme T (2010) Designing ensembles in conformational and sequence space to characterize and engineer proteins. *Curr Opin Struct Biol* 20:377–384.
- Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG (2011) Theoretical and computational protein design. *Annu Rev Phys Chem* 62:129–149.
- Khersonsky O, et al. (2011) Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407:391–412.
- Fleishman SJ, et al. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332:816–821.
- Calhoun JR, et al. (2003) Computational design and characterization of a monomeric helical dinuclear metalloprotein. *J Mol Biol* 334:1101–1115.
- Kono H, Saven JG (2001) Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *J Mol Biol* 306:607–628.
- Butterfoss GL, Kuhlman B (2006) Computer-based design of novel protein structures. *Annu Rev Biophys Biom* 35:49–65.
- Grigoryan G, Keating AE (2008) Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol* 18:477–483.

34. Ghirlanda G, Lear JD, Ogihara NL, Eisenberg D, DeGrado WF (2002) A hierarchic approach to the design of hexameric helical barrels. *J Mol Biol* 319:243–253.
35. Grigoryan G, et al. (2011) Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* 332:1071–1076.
36. North B, Summa CM, Ghirlanda G, DeGrado WF (2001) D(n)-symmetrical tertiary templates for the design of tubular proteins. *J Mol Biol* 311:1081–1090.
37. Gonzalez L, Jr, Brown RA, Richardson D, Alber T (1996) Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nat Struct Biol* 3:1002–1009.
38. Gonzalez L, Jr, Ples JJ, Alber T (1996) An engineered allosteric switch in leucine-zipper oligomerization. *Nat Struct Biol* 3:510–515.
39. Ogihara NL, Weiss MS, DeGrado WF, Eisenberg D (1997) The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies. *Protein Sci* 6:80–88.
40. Wukovitz SV, Yeates TO (1995) Why protein crystals favour some space-groups over others. *Nat Struct Biol* 2:1062–1067.
41. Duan Y, et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24:1999–2012.
42. Fu X, Kono H, Saven JG (2003) Probabilistic approach to the design of symmetric protein quaternary structures. *Protein Eng* 16:971–977.
43. Yang X, Saven JG (2005) Computational methods for protein design and protein sequence variability: Biased Monte Carlo and replica exchange. *Chem Phys Lett* 401:205–210.
44. Kim S, et al. (2005) Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc Natl Acad Sci USA* 102:14278–14283.
45. Senes A, Gerstein M, Engelman DM (2000) Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* 296:921–936.
46. Walters RFS, DeGrado WF (2006) Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci USA* 103:13658–13663.
47. Bender GM, et al. (2007) De novo design of a single-chain diphenylporphyrin metalloprotein. *J Am Chem Soc* 129:10732–10740.
48. Zou JM, Saven JG (2000) Statistical theory of combinatorial libraries of folding proteins: Energetic discrimination of a target structure. *J Mol Biol* 296:281–294.
49. O'Neil KT, DeGrado WF (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250:646–651.
50. Kleiger G, Eisenberg D (2002) GXXXG and GXXXA motifs stabilize FAD and NAD(P)-binding Rossmann folds through C-alpha-H center dot center dot center dot O hydrogen bonds and van der Waals interactions. *J Mol Biol* 323:69–76.
51. Moore DT, Berger BW, DeGrado WF (2008) Protein-protein interactions in the membrane: sequence, structural, and biological motifs. *Structure* 16:991–1001.
52. Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA (1997) Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins* 28:494–514.
53. Juers DH, Matthews BW (2001) Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *J Mol Biol* 311:851–862.
54. Derewenda Z (2004) Rational protein crystallization by mutational surface engineering. *Structure* 12:529–535.
55. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33:491–497.
56. Kantardjieff KA, Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* 12:1865–1871.
57. Derewenda Z, et al. (2007) Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr D* 63:636–645.
58. Grueninger D, et al. (2008) Designed protein-protein association. *Science* 319:206–209.
59. Derewenda Z, Vekilov P (2006) Entropy and surface engineering in protein crystallization. *Acta Crystallogr D* 62:116–124.
60. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.
61. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25:1400–1415.
62. Dunbrack RL, Karplus M (1993) Backbone-dependent rotamer library for proteins—application to side-chain prediction. *J Mol Biol* 230:543–574.